Komisja Egzaminacyjna dla Aktuariuszy

LXXX Egzamin dla Aktuariuszy

Sesja egzaminacyjna w dniu 5 marca 2019 r.

Modelowanie

Imię i nazwisko osoby egzaminowanej: ....

Czas trwania egzaminu: 120 minut

Uwagi

a) W prezentowanych wynikach separatorem dziesiętnym (znakiem dziesiętnym) jest kropka „ .”.

b) W prezentowanych wynikach oszacowań uogólnionych modeli liniowych (GLM):
- Residual deviance i Resid. Dev – oznacza dewiancję oszacowanego modelu,
- Null deviance – oznacza dewiancję modelu zerowego,
- Deviance – redukcję dewiancji po dodaniu kolejnej zmiennej objaśniającej,
- Df – stopnie swobody,
- Sum Sq – suma kwadratów.

c) W zadaniach wartość zagrożona na poziomie ufności α jest definiowana jako kwantyl rzędu α rozkładu odpowiedniej zmiennej losowej, tzn.

$$VaR_\alpha(X) = \inf\{x: F_X(x) \geq \alpha\} .$$

d) W zadaniach zastosowano następujące oznaczenia:
$E(X)$ – wartość oczekiwana
$D(X)$ – odchylenie standardowe

e) Wartości $\chi^2_{\alpha;\upsilon}$ rozkładu chi-kwadrat spełniające warunek $P(\chi^2 \geq \chi^2_{\alpha;\upsilon}) = \alpha$

υ \ α	**0.995**	**0.99**	**0.975**	**0.95**	**0.9**	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801

f) Wybrane kwantyle standardowego rozkładu normalnego

α	**0.900**	**0.950**	**0.975**	**0.985**	**0.989**	**0.990**	**0.991**	**0.995**	**0.998**	**0.999**
u_α	1.282	1.645	1.960	2.170	2.290	2.326	2.366	2.576	2.878	3.090

g) Wartości dystrybuanty w punkcie 1000 (tzn. $F(1000)$) dla rozkładu $Gamma(\alpha,\lambda)$ (dla wybranych parametrów α,λ).
Uwaga! Przyjęto parametryzację, w której $E(X)=\frac{\alpha}{\lambda}$, $D^2(X)=\frac{\alpha}{\lambda^2}$

α \ λ	***0,00059***	***0,00060***	***0,00061***	***0,00062***	***0,00063***	***0,00064***	***0,00065***	***0,00066***	***0,00067***	***0,00068***
0,485	0.7315	0.7355	0.7394	0.7432	0.7469	0.7506	0.7542	0.7577	0.7612	0.7646
0,486	0.7310	0.7349	0.7388	0.7426	0.7463	0.7500	0.7536	0.7571	0.7606	0.7640
0,487	0.7304	0.7343	0.7382	0.7420	0.7458	0.7494	0.7530	0.7566	0.7601	0.7635
0,488	0.7298	0.7337	0.7376	0.7414	0.7452	0.7489	0.7525	0.7560	0.7595	0.7630
0,489	0.7292	0.7331	0.7370	0.7409	0.7446	0.7483	0.7519	0.7555	0.7590	0.7624
0,490	0.7286	0.7326	0.7365	0.7403	0.7440	0.7477	0.7514	0.7549	0.7584	0.7619
0,491	0.7280	0.7320	0.7359	0.7397	0.7435	0.7472	0.7508	0.7544	0.7579	0.7613
0,492	0.7274	0.7314	0.7353	0.7391	0.7429	0.7466	0.7503	0.7538	0.7573	0.7608
0,493	0.7268	0.7308	0.7347	0.7386	0.7423	0.7461	0.7497	0.7533	0.7568	0.7603
0,494	0.7262	0.7302	0.7341	0.7380	0.7418	0.7455	0.7491	0.7527	0.7563	0.7597
0,495	0.7256	0.7296	0.7335	0.7374	0.7412	0.7449	0.7486	0.7522	0.7557	0.7592
0,496	0.7250	0.7290	0.7330	0.7368	0.7406	0.7444	0.7480	0.7516	0.7552	0.7586
0,497	0.7244	0.7284	0.7324	0.7363	0.7401	0.7438	0.7475	0.7511	0.7546	0.7581
0,498	0.7238	0.7279	0.7318	0.7357	0.7395	0.7432	0.7469	0.7505	0.7541	0.7575
0,499	0.7232	0.7273	0.7312	0.7351	0.7389	0.7427	0.7463	0.7500	0.7535	0.7570
0,500	0.7226	0.7267	0.7306	0.7345	0.7383	0.7421	0.7458	0.7494	0.7530	0.7565
0,501	0.7221	0.7261	0.7301	0.7340	0.7378	0.7415	0.7452	0.7489	0.7524	0.7559
0,502	0.7215	0.7255	0.7295	0.7334	0.7372	0.7410	0.7447	0.7483	0.7519	0.7554
0,503	0.7209	0.7249	0.7289	0.7328	0.7366	0.7404	0.7441	0.7477	0.7513	0.7548
0,504	0.7203	0.7243	0.7283	0.7322	0.7361	0.7398	0.7436	0.7472	0.7508	0.7543
0,505	0.7197	0.7237	0.7277	0.7317	0.7355	0.7393	0.7430	0.7466	0.7502	0.7538
0,506	0.7191	0.7232	0.7272	0.7311	0.7349	0.7387	0.7424	0.7461	0.7497	0.7532
0,507	0.7185	0.7226	0.7266	0.7305	0.7344	0.7381	0.7419	0.7455	0.7491	0.7527
0,508	0.7179	0.7220	0.7260	0.7299	0.7338	0.7376	0.7413	0.7450	0.7486	0.7521
0,509	0.7173	0.7214	0.7254	0.7293	0.7332	0.7370	0.7408	0.7444	0.7480	0.7516
0,510	0.7167	0.7208	0.7248	0.7288	0.7326	0.7365	0.7402	0.7439	0.7475	0.7510
0,511	0.7161	0.7202	0.7242	0.7282	0.7321	0.7359	0.7396	0.7433	0.7469	0.7505
0,512	0.7155	0.7196	0.7237	0.7276	0.7315	0.7353	0.7391	0.7428	0.7464	0.7500
0,513	0.7149	0.7190	0.7231	0.7270	0.7309	0.7348	0.7385	0.7422	0.7458	0.7494
0,514	0.7143	0.7185	0.7225	0.7265	0.7304	0.7342	0.7380	0.7417	0.7453	0.7489
0,515	0.7138	0.7179	0.7219	0.7259	0.7298	0.7336	0.7374	0.7411	0.7447	0.7483
0,516	0.7132	0.7173	0.7213	0.7253	0.7292	0.7331	0.7368	0.7406	0.7442	0.7478
0,517	0.7126	0.7167	0.7208	0.7247	0.7287	0.7325	0.7363	0.7400	0.7436	0.7472
0,518	0.7120	0.7161	0.7202	0.7242	0.7281	0.7319	0.7357	0.7394	0.7431	0.7467
0,519	0.7114	0.7155	0.7196	0.7236	0.7275	0.7314	0.7352	0.7389	0.7426	0.7462
0,520	0.7108	0.7149	0.7190	0.7230	0.7269	0.7308	0.7346	0.7383	0.7420	0.7456

h) Wartości F_{r_1,r_2} rozkładu F spełniające warunek $P(F \geq F_{r_1,r_2}) = 0.05$

r_2 \ r_1	***1***	***2***	***3***	***4***	***5***	***6***	***7***
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103
700	3.855	3.009	2.618	2.385	2.227	2.112	2.023
800	3.853	3.007	2.616	2.383	2.225	2.110	2.021
900	3.852	3.006	2.615	2.382	2.224	2.109	2.020
1000	3.851	3.005	2.614	2.381	2.223	2.108	2.019
∞	3.844	2.998	2.607	2.374	2.216	2.101	2.012

r_2 \ r_1	***8***	***9***	***10***	***11***	***12***	***13***	***14***
1	238.883	240.543	241.882	242.983	243.906	244.690	245.364
2	19.371	19.385	19.396	19.405	19.413	19.419	19.424
3	8.845	8.812	8.786	8.763	8.745	8.729	8.715
4	6.041	5.999	5.964	5.936	5.912	5.891	5.873
5	4.818	4.772	4.735	4.704	4.678	4.655	4.636
6	4.147	4.099	4.060	4.027	4.000	3.976	3.956
7	3.726	3.677	3.637	3.603	3.575	3.550	3.529
8	3.438	3.388	3.347	3.313	3.284	3.259	3.237
9	3.230	3.179	3.137	3.102	3.073	3.048	3.025
10	3.072	3.020	2.978	2.943	2.913	2.887	2.865
100	2.032	1.975	1.927	1.886	1.850	1.819	1.792
700	1.952	1.893	1.844	1.802	1.766	1.734	1.706
800	1.950	1.892	1.843	1.801	1.764	1.732	1.704
900	1.949	1.890	1.841	1.799	1.763	1.731	1.703
1000	1.948	1.889	1.840	1.798	1.762	1.730	1.702
∞	1.941	1.882	1.833	1.791	1.755	1.723	1.694

Zadanie 1.

Liczbę szkód *K* zgłaszanych przez kierowców z pewnego portfela ubezpieczeń AC modelowano z wykorzystaniem regresji Poissona (uogólnionego modelu liniowego dla zmiennej zależnej o rozkładzie Poissona), uwzględniając następujące zmienne objaśniające:

Zmienna	*Opis*
region	Region zamieszkania kierowcy. Zmienna jakościowa przyjmująca następujące kategorie: *R0, R1, R2, R3, R4, R5, R6, R7, R8.*
klasa.ryzyka	Zmienna jakościowa określająca klasę ryzyka, którą przypisano kierowcy. Przyjmuje następujące kategorie: *kl.1, kl.2, kl.3, kl.4, kl.5 i kl.6* (im wyższa klasa, tym większa zniżka za bezszkodowość).
prawo.jazdy	Zmienna jakościowa określająca doświadczenie kierowcy. Przyjmuje następujące kategorie: *1, 2, 3, 4, 5, 6, 7 i 8+* (poszczególne kategorie oznaczają ile lat kierowca posiada prawo jazdy).
prior.szkody	Liczba wcześniejszych szkód (zmienna ilościowa).
wiek.kierowcy	Zmienna jakościowa określająca wiek kierowcy. Przyjmuje następujące kategorie: *18, 19, 20, 21, 22, 23, 24, 25, 26+* (poszczególne kategorie oznaczają wiek kierowcy).
polisa	Zmienna jakościowa określająca, czy polisa jest nowa, czy odnowiona. Przyjmuje następujące kategorie: *np, op.*
stan.cywilny	Zmienna jakościowa określająca stan cywilny kierowcy. Przyjmuje następujące kategorie: *M* – małżeństwo, *S* – singiel, *R* – rozwiedziony/rozwiedziona, *W* – wdowiec/wdowa).
plec	Zmienna jakościowa określająca płeć kierowcy. Przyjmuje następujące kategorie: *K* - kobieta, *M* – mężczyzna.

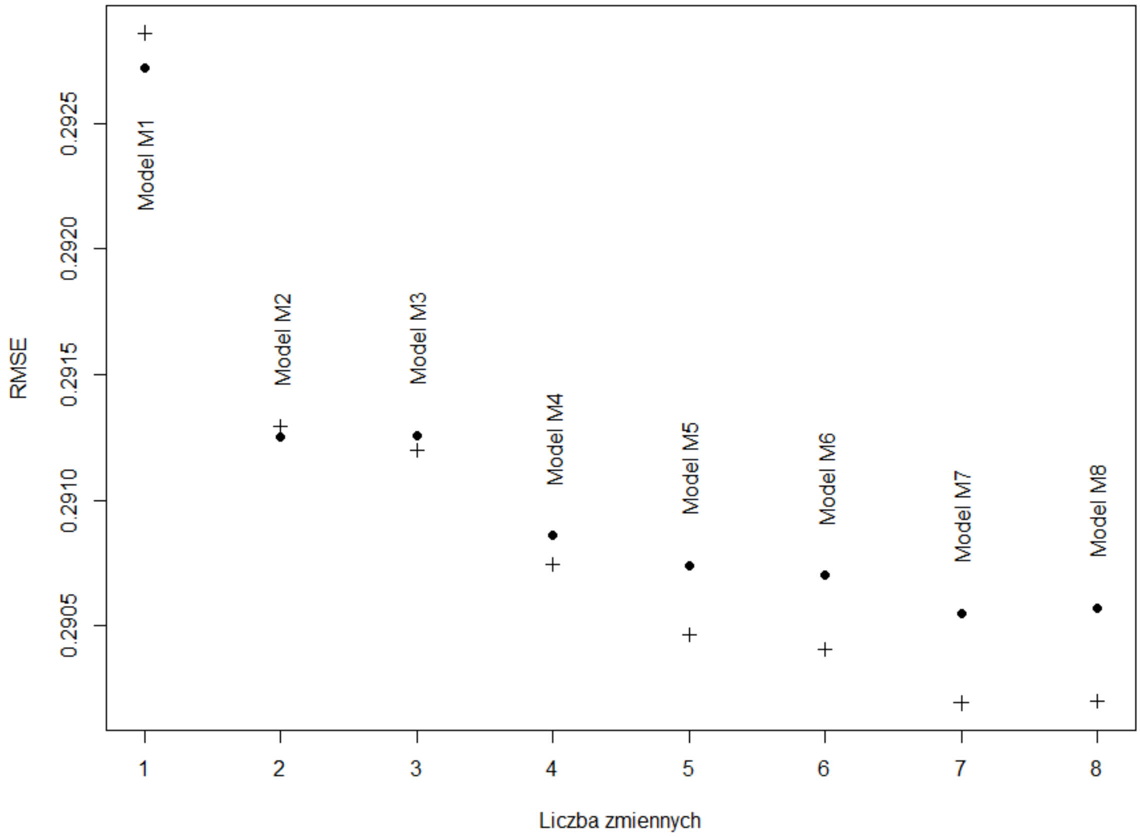
Oszacowano 8 modeli, przy czym w każdym kolejnym uwzględniano jedną dodatkową zmienną objaśniającą w stosunku do poprzedniego (por. Tab.1.):

Tabela 1.

Model	*Zmienne objaśniające*
M1	*region*
M2	*region + klasa.ryzyka*
M3	*region + klasa.ryzyka + prawo.jazdy*
M4	*region + klasa.ryzyka + prawo.jazdy + prior.szkody*
M5	*region + klasa.ryzyka + prawo.jazdy + prior.szkody + wiek.kierowcy*
M6	*region + klasa.ryzyka + prawo.jazdy + prior.szkody + wiek.kierowcy + polisa*
M7	*region + klasa.ryzyka + prawo.jazdy + prior.szkody + wiek.kierowcy + polisa + stan.cywilny*
M8	*region + klasa.ryzyka + prawo.jazdy + prior.szkody + wiek.kierowcy + polisa + stan.cywilny + plec*

Na poniższym rysunku przedstawiono wyniki *10*-krotnej walidacji krzyżowej tych modeli. Symbol „+" oznacza średniokwadratowy błąd predykcji ex post (*root mean square error,* RMSE) dla zbioru uczącego, natomiast symbol „●" - średni RMSE otrzymany z *10*-krotnej walidacji krzyżowej (na zbiorze uczącym).

Rysunek 1



a) Jakie typowe zjawisko, które może wystąpić podczas modelowania ilustruje rysunek 1? Nazwać i omówić to zjawisko.
b) W oparciu o podane informacje wybrać najlepszy model. Wybór uzasadnić.

Odpowiedzi:

……………………………………………………………………………………………………

Odp. a)

W odpowiedzi należało uwzględnić następujące elementy:

- Nazwę zjawiska: nadmierne dopasowanie (przeuczenie, przetrenowanie, overfitting) modelu.
- Opis tego zjawiska (wytłumaczenie na czym polega): Na przykład: Z nadmiernym dopasowaniem mamy do czynienia, gdy model statystyczny ma zbyt dużo parametrów w stosunku do rozmiaru próby na podstawie której był konstruowany. Taki model jest bardzo dobrze dopasowany do danych uczących, jednak daje gorsze wyniki, gdy zastosujemy go do danych, z którymi się „nie zetknął" podczas uczenia.
- Wytłumaczenie dlaczego przedstawiony rysunek ilustruje to zjawisko.

Odp. b)

W odpowiedzi należało uwzględnić następujące elementy:

- Wskazanie modelu: M4 (Można było wskazać także M7 i ew. M5.)
- Uzasadnienie, zawierające omówienie różnicy poziomu błędów RMSE dla zbioru uczącego i zbiorów testowych (wynikających z walidacji krzyżowej) w kontekście wzrastającej liczby parametrów kolejnych modeli. Na przykład: Model M4 w porównaniu z poprzednimi (M1, M2, M3) ma zauważalnie niższy poziom błędu RMSE, zarówno na danych treningowych jak i testowych. W kolejnych modelach (M5-M8) można zauważyć znaczący przyrost liczby parametrów, który skutkuje jednak zauważalnym obniżeniem błędu RMSE tylko na zbiorze uczącym.

Uwaga! Dopuszczalne było wskazanie innych modeli (oprócz M1) z „rozsądnym” uzasadnieniem.

Rozwiązanie:

Zadanie 2.

W trakcie konstrukcji modelu wysokości pojedynczego roszczenia (*severity model*) dla pewnego portfela ubezpieczeń AC brano pod uwagę następujące zmienne objaśniające:

Zmienna	*Opis*
wiek.sam	Wiek samochodu w latach (zmienna ilościowa).
typ.nadwozia	Zmienna jakościowa określająca typ nadwozia samochodu. Przyjmuje następujące kategorie: *A, B, C, D, E, F, G, H.*
paliwo	Zmienna jakościowa określająca rodzaj paliwa. Przyjmuje następujące kategorie: *B* – benzyna, *D* – olej napędowy, *LPG* – gaz).
moc	Moc silnika w KM (zmienna ilościowa).
polisa	Zmienna jakościowa określająca, czy polisa jest nowa, czy odnowiona. Przyjmuje następujące kategorie: *np, op.*
stan.cywilny	Zmienna jakościowa określająca stan cywilny kierowcy. Przyjmuje następujące kategorie: *M* – małżeństwo, *S* – singiel, *R* – rozwiedziony/rozwiedziona, *W* – wdowiec/wdowa).
plec	Zmienna jakościowa określająca płeć kierowcy. Przyjmuje następujące kategorie: *K* - kobieta, *M* – mężczyzna.
prawo.jazdy	Zmienna jakościowa określająca doświadczenie kierowcy. Przyjmuje następujące kategorie: *1, 2, 3, 4, 5, 6, 7 i 8+* (poszczególne kategorie oznaczają ile lat kierowca posiada prawo jazdy).
klasa.ryzyka	Zmienna jakościowa określająca klasę ryzyka, którą przypisano kierowcy. Przyjmuje następujące kategorie: *kl.1, kl.2, kl.3, kl.4, kl.5 i kl.6* (im wyższa klasa, tym większa zniżka za bezszkodowość).

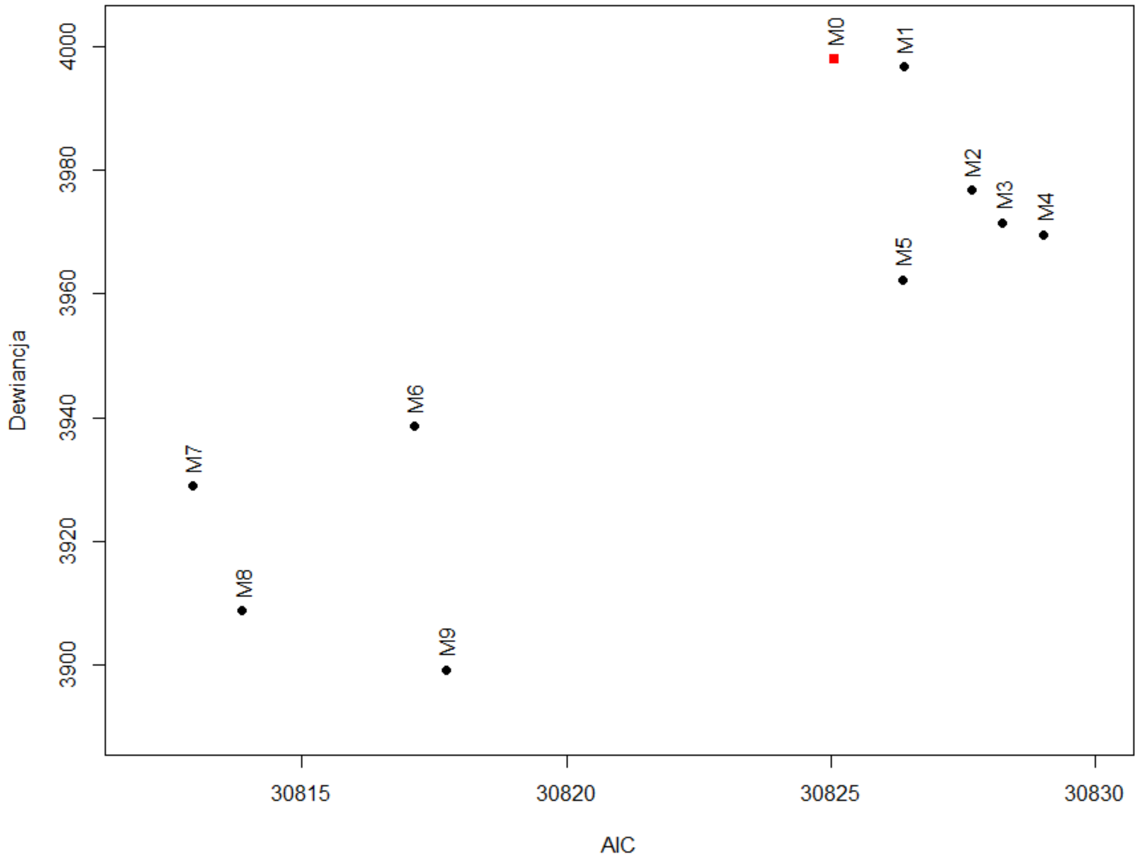
Wykorzystując zbiór uczący liczący 1856 obserwacji, oszacowano 10 uogólnionych modeli liniowych (por. Tab.2.). W każdym z nich przyjęto **rozkład gamma** dla zmiennej objaśnianej oraz **logarytmiczną funkcję wiążącą**.

Tabela 2

Model	*Zmienne objaśniające*	*AIC*	*Dewiancja*
M0	*bez zmiennych objaśniających (model zerowy)*	30825.07	3997.821
M1	*wiek.sam*	30826.38	3996.731
M2	*wiek.sam + typ.nadwozia*	30827.66	3976.792
M3	*wiek.sam + typ.nadwozia + paliwo*	30828.25	3971.468
M4	*wiek.sam + typ.nadwozia + paliwo + moc*	30829.02	3969.543
M5	*wiek.sam + typ.nadwozia + paliwo + moc + polisa*	30826.35	3962.252
M6	*wiek.sam + typ.nadwozia + paliwo + moc + polisa + stan.cywilny*	30817.12	3938.545
M7	*wiek.sam + typ.nadwozia + paliwo + moc + polisa + stan.cywilny + plec*	30812.93	3928.949
M8	*wiek.sam + typ.nadwozia + paliwo + moc + polisa + stan.cywilny + plec + prawo.jazdy*	30813.86	3908.763
M9	*wiek.sam + typ.nadwozia + paliwo + moc + polisa + stan.cywilny + plec + prawo.jazdy + klasa.ryzyka*	30817.72	3899.306

Na poniższym rysunku (Rys. 2.) przedstawiono wartości kryterium informacyjnego Akaike'a i dewiancje oszacowanych modeli.

Rysunek 2.



Parametr dyspersji dla modelu M6 wynosi: 2.0038; dla M7: 1.9825; dla M8: 1.9498; dla M9: 1.9671.

Na podstawie podanych wyników:

a) Scharakteryzować widoczne dwie grupy modeli.
b) Spośród modeli od M6 do M9 wybrać najlepszy. Wybór uzasadnić (w uzasadnieniu należy **między innymi** posłużyć się odpowiednim testem).

Odpowiedzi:

....

Odp. a)

W odpowiedzi należało uwzględnić następujące elementy:

- Wskazanie modeli należących do dwóch grup: Jedną grupę stanowią modele M0-M5, drugą M6-M9.
- Charakterystykę grup, w której zwraca się uwagę na:
  - odległość modeli z poszczególnych grup od modelu zerowego (bez predyktorów) ze względu na miernik *AIC* i dewiancję,
  - to, że znaczną poprawę w sensie mniejszej dewiancji i mniejszego AIC uzyskuje się po uwzględnieniu w modelu grupy zmiennych charakteryzujących kierowców.

....

Odp. b)

W odpowiedzi należało uwzględnić następujące elementy:

- Wskazanie najlepszego modelu: M8 (ew. M7, M9).
- Uzasadnienie, w którym:
  - zwraca się uwagę na odległość analizowanych modeli od modelu zerowego M0,
  - porównuje się wzrost AIC po uwzględnieniu kolejnych zmiennych objaśniających,
  - porównuje się istotność obniżenia dwiancji po uwzględnieniu kolejnych zmiennych objaśniających z wykorzystaniem testu F (ze względu na oszacowania parametru dyspersji), którego statystyka wyraża się wzorem:

$$F = \frac{D\left(y; \hat{\theta}^P\right) - D\left(y; \hat{\theta}^Q\right)}{\hat{\phi}(q-p)},$$

gdzie $\hat{\phi}$ oszacowanie parametru dyspersji.

Uwaga ! Akceptowane było także wykorzystanie testu ilorazu wiarygodności.

Rozwiązanie:

Zadanie 3.

W poniższej tabeli (Tab. 3.) podano oszacowania parametrów modeli M0, M6, M7, M8 i M9 przedstawionych w poprzednim zadaniu (**Zadanie 2**). Szacując parametry przyjęto jednostkowe wagi dla każdej obserwacji.

Tabela 3

	M0	**M6**	**M7**	**M8**	**M9**
Wyraz wolny	6.6694	6.4492	6.4234	6.4521	6.4405
wiek.sam		0.0103	0.0092	0.0095	0.0138
*typ.nadwozia*_B		0.4142	0.4354	0.4191	0.4233
*typ.nadwozia*_C		-0.0633	-0.0532	-0.0315	-0.0497
*typ.nadwozia*_D		0.0663	0.0633	0.0512	0.0572
*typ.nadwozia*_E		0.0932	0.0956	0.1184	0.1290
*typ.nadwozia*_F		-1.0055	-0.9829	-0.9459	-0.9603
*typ.nadwozia*_G		0.2219	0.2241	0.2349	0.2607
*typ.nadwozia*_H		-0.0818	-0.0986	-0.1063	-0.0547
*paliwo*_B		-0.5910	-0.5571	-0.5760	-0.5898
*paliwo*_LPG		-0.6654	-0.7045	-0.7254	-0.7388
moc		0.0019	0.0019	0.0018	0.0020
polisa_op		-0.1375	-0.1287	-0.0945	-0.0524
stan.cywilny_R		-0.0304	-0.0649	-0.0710	-0.0976
stan.cywilny_S		0.5045	0.4867	0.4730	0.4506
stan.cywilny_W		-0.0600	-0.0696	-0.0897	-0.0883
plec_K			0.2054	0.2047	0.2075
prawo.jazdy_2				-0.0330	0.0210
prawo.jazdy_3				0.0008	0.0770
prawo.jazdy_4				-0.0729	0.0210
prawo.jazdy_5				-0.0466	0.0796
prawo.jazdy_6				-0.5679	-0.4383
prawo.jazdy_7				0.0341	0.1873
prawo.jazdy_8+				0.1379	0.2808
klasa.ryzyka_ kl.2					-0.2494
klasa.ryzyka_ kl.3					-0.1088
klasa.ryzyka_ kl.4					-0.1510
klasa.ryzyka_ kl.5					-0.2141
klasa.ryzyka_ kl.6					-0.2475
Parametr dyspersji	2.0492	2.0038	1.9825	1.9498	1.9671

a) Dla wybranego w poprzednim zadaniu (**Zadanie 2**) modelu określić bazową grupę kierowców.
b) Wyznaczyć prawdopodobieństwo, że wysokość pojedynczego roszczenia zgłoszonego przez kierowcę z grupy bazowej posiadającego 5-letni samochód o mocy 100 KM będzie większa od 1000.
c) Dla tego kierowcy (tzn. z grupy bazowej i posiadającego 5-letni samochód o mocy 100 KM) wyznaczyć składkę zgodnie z zasadą odchylenia standardowego. Przyjąć parametr bezpieczeństwa $\delta = 1$ i założyć, że roczna liczba szkód ma rozkład Poissona ze średnią 0.1.

Uwaga!! W przypadku nierozwiązania zadania 2 powyższe polecenia a), b) i c) dotyczą modelu M9.

Odpowiedzi:

Odp. a)

Bazowa grupa kierowców:
typ.nadwozia: A
paliwo: *D* – olej napędowy
polisa: *np* - nowa polisa
stan.cywilny: *M* – małżeństwo
plec: M - meżczyna
prawo.jazdy: *1* - kierowca posiada prawo jazdy przez 1 rok
klasa.ryzyka: *kl.1* – pierwsza klasa ryzyka

Model M6
Model M7
Model M8
Model M9

Odp. b)

Wariancja: $\hat{\phi}\hat{\mu}_i^2$

Parametry rozkładu gamma: $E(X) = \frac{\alpha}{\lambda}$, $D^2(X) = \frac{\alpha}{\lambda^2}$ stąd $\alpha = \frac{(E(X))^2}{D^2(X)}$, $\lambda = \frac{E(X)}{D^2(X)}$

Model:	***M6***	***M7***	***M8***	***M9***
Predyktor	6.6907	6.6594	6.6796	6.7095
Średnia	804.8855	780.0827	796.0006	820.1605
Wariancja	1298143.05	1206408.93	1235426.49	1323195.73
α	0.4991	0.5044	0.5129	0.5084
λ	0.00062	0.00065	0.00064	0.00062
$P(Y \leq 1000)$	0,7351	0,7421	0,7365	0,7296
$P(Y > 1000)$	**0,2649**	**0,2579**	**0,2635**	**0,2704**

Odp. c)

Składka: $E(Z) + \delta D(Z)$

$$E(Z) = E(K) \cdot E(X) = \mu_K \mu_X$$
$$Var(Z) = Var(K) \cdot (E(X))^2 + E(K) \cdot Var(X) = \sigma_K^2 (\mu_X)^2 + \mu_K \sigma_X^2$$

Model:	***M6***	***M7***	***M8***	***M9***
$E(Z)$	80.4886	78.0083	79.6000	82.0160
$Var(Z)$	194598.37	181493.80	186904.35	199585.89
$D(Z)$	441.1330	426.0209	432.3244	446.7504
Składka	521.6216	504.0292	511.9244	528.7764

Rozwiązanie:

Zadanie 4.

Aktuariusz analizował dwa rodzaje ryzyka X_1 i X_2. Dopasował rozkłady brzegowe, a ponieważ ich charakter wskazywał na możliwość jednoczesnego wystąpienia dużych strat dla X_1 i X_2 , do modelowania struktury zależności między nimi wybrał kopulę Claytona:

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}.$$

Następnie wyznaczył, podane w poniższej tabeli, kwantyle rozkładów brzegowych i wartości kopuli $C(u,u)$ dla u malejącego od 0.99 do 0.01.

u	$F_{X_1}^{-1}(u)$	$F_{X_2}^{-1}(u)$	$C(u,u)$
0.01	-232,63	-465,27	0.00794
0.02	-205,37	-410,75	0.01587
0.03	-188,08	-376,16	0.02381
0.04	-175,07	-350,14	0.03175
0.05	-164,49	-328,97	0.03969
0.06	-155,48	-310,95	0.04762
0.07	-147,58	-295,16	0.05556
0.08	-140,51	-281,01	0.06350
0.09	-134,08	-268,15	0.07144
0.10	-128,16	-256,31	0.07938
⋮	⋮	⋮	⋮
0.90	128,16	256,31	0.83086
0.91	134,08	268,15	0.84558
0.92	140,51	281,01	0.86069
0.93	147,58	295,16	0.87623
0.94	155,48	310,95	0.89222
0.95	164,49	328,97	0.90870
0.96	175,07	350,14	0.92572
0.97	188,08	376,16	0.94330
0.98	205,37	410,75	0.96151
0.99	232,63	465,27	0.98039

a) Krótko omówić problem pomiaru i modelowania zależności między dwoma rodzajami ryzyka, które mogą charakteryzować się możliwością jednoczesnego wystąpienia strat ekstremalnych. Wskazać odpowiednie narzędzia. Czy takim narzędziem może być współczynnik korelacji liniowej Pearsona?

b) Wykorzystując współczynnik zależności w dolnym ogonie (*coefficient of lower tail dependence*) λ_L i informacje z tabeli ustalić parametr kopuli Claytona, którą wykorzystał aktuariusz do modelowania zależności między X_1 i X_2.

Odpowiedzi:

....

Odp. a)

W odpowiedzi należało uwzględnić następujące elementy:

- Wskazanie, że najbardziej popularna miara zależności, czyli współczynnik korelacji liniowej nie może być wykorzystany, gdyż można go stosować, gdy X_1 i X_2 mają dwuwymiarowy rozkład normalny, czyli brzegowe muszą być normalne. W przypadku występowania wartości ekstremalnych nie można tego założyć.
- Wskazanie odpowiednich narzędzi pomiaru zależności: współczynniki korelacji rang (rang Spearmana, tau Kendalla), współczynniki zależności w dolnym i górnym ogonie.
- Wskazanie odpowiednich narzędzi modelowania zależności: np. kopule.

....

Odp. b)

Współczynnik zależności w dolnym ogonie: $\lambda_L = \lim_{u \to 0} \frac{C(u,u)}{u}$

Dla kopuli Claytona : $\lambda_L = \lim_{u \to 0} \frac{C(u,u)}{u} = \lim_{u \to 0} \frac{\left(u^{-\theta}+u^{-\theta}-1\right)^{-\frac{1}{\theta}}}{u} = 2^{-\frac{1}{\theta}}$

Wykorzystując dane z tabeli otrzymujemy:

$2^{-\frac{1}{\theta}} \approx \frac{C(0.01,0.01)}{0.01} = 0.794$ stąd $\theta = 3.0049$

Uwaga! Akceptowane było także inne rozwiązanie.

Rozwiązanie:

Zadanie 5.

Przedłużenie przez kierowcę ubezpieczenia AC w firmie XYZ modelowano z wykorzystaniem regresji logistycznej. Przyjęto, że zmienna zależna Y, przyjmuje dwie wartości: $Y = 1$, gdy kierowca przedłuży umowę na kolejny rok oraz $Y = 0$, gdy nie przedłuży. Uwzględniono następujące zmienne objaśniające:

- *plec:* Płeć (K - kobieta, M – mężczyzna)
- *zamieszkanie:* Miejsce zamieszkania (Miasto – miasto, Wies – wieś)
- *stan.cywilny:* Stan cywilny (C – małżeństwo, S – singiel, O – inny)
- *wiek:* Zmienna jakościowa przyjmująca trzy kategorie: W1 (młodzi kierowcy), W2 (doświadczeni kierowcy), W3 (starsi kierowcy).
- *lojalnosc:* Liczba lat, w których kierowca był klientem firmy XYZ.

Na podstawie zbioru uczącego liczącego 3109 obserwacji, metodą największej wiarygodności oszacowano następujący model:

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	1.1116	0.2559	0.0000
*plec*_M	-1.0507	0.0961	< 2e-16
*zamieszkanie*_Wies	-1.1027	0.0871	< 2e-16
*stan.cywilny*_C	-0.3720	0.1060	0.0005
*stan.cywilny*_O	-0.1561	0.2253	0.4884
*wiek*_W2	-1.1991	0.2587	0.0000
*wiek*_W3	-2.3648	0.2956	0.0000
lojalnosc	0.1120	0.0071	< 2e-16

Null deviance: 3967 on 3108 degrees of freedom
Residual deviance: 3385 on 3101 degrees of freedom
AIC: 3401

Dla tego modelu tabela analizy dewiancji (*Deviance Table*) przedstawia się następująco:

Model: binomial, link: logit
Response: Y
Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			3108	3967,0	
plec	1	107,762	3107	3859,3	< 2,2e-16
zamieszkanie	1	136,760	3106	3722,5	< 2,2e-16
stan.cywilny	2	21,619	3104	3700,9	2,02e-05
wiek	2	36,709	3102	3664,2	1,07e-08
lojalnosc	1	279,150	3101	3385,0	< 2,2e-16

Na zbiorze testowym liczącym 100 obserwacji ustalono, że przy założeniu równych kosztów błędnych klasyfikacji, optymalny punkt odcięcia jest równy oszacowanemu prawdopodobieństwu przedłużenia umowy dla następującego kierowcy:

- *plec:* M (mężczyzna),
- *zamieszkanie:* Miasto (miasto),
- *stan.cywilny:* C (żonaty),
- *wiek:* W3,
- *lojalnosc:* 14 lat.

Dla tego punktu odcięcia, **bez uwzględnionych dwóch obserwacji ze zbioru testowego** podanych w poniższej tabeli:

Zmienna	Wartości zmiennych objaśniających odpowiadające nieuwzględnionej obserwacji	
	pierwszej	drugiej
plec	M	M
zamieszkanie	Wies	Wies
stan.cywilny	C	C
wiek	W3	W3
lojalnosc	26	18
Y	0	0

otrzymano następującą tabelę trafności prognoz:

		Stan faktyczny	
		$Y_i = 0$ (N)	$Y_i = 1$ (P)
Prognoza	$Y_i^P = 0$ (N)	48	5
	$Y_i^P = 1$ (P)	15	30

a) Podać definicje i interpretacje specyficzności (*specificity*) i czułości (s*ensitivity*). W jaki sposób wykorzystuje się te miary do ustalenia optymalnego punktu odcięcia?

b) Dla optymalnego punktu odcięcia (który wskazano w treści zadania) obliczyć specyficzność i czułość, wykorzystując wszystkie obserwacje ze zbioru testowego.

Odpowiedzi:

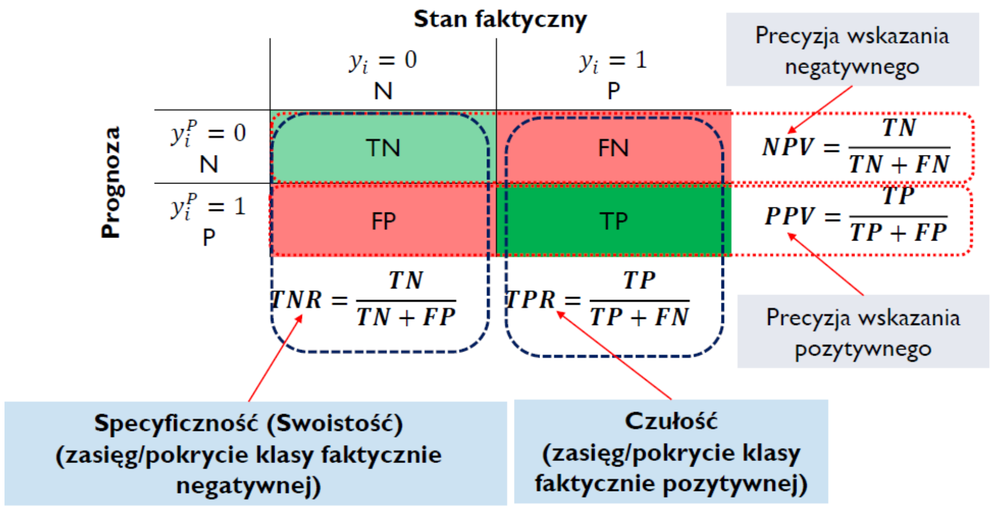
....

Odp. a)

W odpowiedzi należało uwzględnić następujące elementy:

- Definicję i interpretację specyficzności.
- Definicję i interpretację czułości.
- Wskazanie, w jaki sposób wykorzystuje się te miary do ustalenia optymalnego punktu odcięcia.

Tabela trafności prognoz:



....

Odp. b)

Predyktor: $logit_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$

Prawdopodobieństwo: $\hat{p}_i = \frac{\exp(logit_i)}{1+\exp(logit_i)}$

Prognoza: $y_i^P = \begin{cases} 0, & gdy\, \hat{p}_i < p_{pr} \\ 1, & gdy\, \hat{p}_i \geq p_{pr} \end{cases},$

gdzie p_{pr} jest wartością progową (decyzyjną, punktem odcięcia).

Punkt odcięcia p_{pr}:

logit: $logit = 1.1116 - 1.0507 - 0.3720 - 2.3648 + 0.1120 \cdot 14 = -1{,}1079.$

Prawdopodobieństwo: $\hat{p} = \frac{\exp(-1{,}1079)}{1+\exp(-1{,}1079)} = 0.24826$

Punkt odcięcia: $p_{pr} = 0.24826$

Prognoza:

- Pierwsza obserwacja:
  logit: $logit =$ -0,8666
  Prawdopodobieństwo: 0,29596
  Prognoza: $Y_i^P = 1$

- Druga obserwacja:
  logit: $logit =$ -1,7626
  Prawdopodobieństwo: 0,14647
  Prognoza: $Y_i^P = 0$

Tabela trafności prognoz (po uwzględnieniu wszystkich obserwacji ze zbioru testowego):

		Stan faktyczny	
		$Y_i = 0$ (N)	$Y_i = 1$ (P)
Prognoza	$Y_i^P = 0$ (N)	49	5
	$Y_i^P = 1$ (P)	16	30

Specyficzność: 0.7538462

Czułość: 0.8571429

Rozwiązanie:

Zadanie 6.

Zakład ubezpieczeń posiada dwa segmenty ubezpieczeń innych niż ubezpieczenia na życie (segment s i t). W poniższej tabeli podano następujące informacje dotyczące tych segmentów:

– miarę wielkości ryzyka składki ($V_{(prem,i)}, i = s, t$),
– miarę wielkości ryzyka rezerw ($V_{(res,s)}, i = s, t$),
– odchylenie standardowe ryzyka składki ($\sigma_{(prem,i)}, i = s, t$),
– odchylenie standardowe ryzyka rezerw($\sigma_{(res,i)}, i = s, t$),

oraz parametr zależności ryzyka składki i rezerw między tymi segmentami ($CorrS_{s,t}$).

Segment *s*	**Segment *t***
$V_{(prem,s)} = 100$	$V_{(prem,t)} = 100$
$V_{(res,s)} = 100$	$V_{(res,t)} = 100$
$\sigma_{(prem,s)} = 0.10$	$\sigma_{(prem,t)} = 0.08$
$\sigma_{(res,s)} = 0.09$	$\sigma_{(res,t)} = 0.08$
$CorrS_{s,t} = 0.5$	

Ponadto wiadomo, że dla tego zakładu ubezpieczeń:

– kapitałowy wymóg wypłacalności dla podmoduł ryzyka katastroficznego w ubezpieczeniach innych niż ubezpieczenia na życie wynosi: 100,
– kapitałowy wymóg wypłacalności dla podmodułu ryzyka związanego z rezygnacjami w ubezpieczeniach innych niż ubezpieczenia na życie wynosi: 10.

Parametry zależności ryzyka aktuarialnego w ubezpieczeniach innych niż ubezpieczenia na życie przedstawiają się następująco:

	Ryzyko składki i rezerw w ubezpieczeniach innych niż ubezpieczenia na życie	Ryzyko katastroficzne w ubezpieczeniach innych niż ubezpieczenia na życie	Ryzyko związane z rezygnacjami w ubezpieczeniach innych niż ubezpieczenia na życie
Ryzyko składki i rezerw w ubezpieczeniach innych niż ubezpieczenia na życie	1	0.25	0
Ryzyko katastroficzne w ubezpieczeniach innych niż ubezpieczenia na życie	0.25	1	0
Ryzyko związane z rezygnacjami w ubezpieczeniach innych niż ubezpieczenia na życie	0	0	1

a) Wyznaczyć wymóg kapitałowy dla ryzyka aktuarialnego w ubezpieczeniach innych niż ubezpieczenia na życie dla tego zakładu. Skorzystać ze standardowej formuły Solvency II.
b) Wyznaczyć efekt dywersyfikacji dla tego ryzyka (tzn. ryzyka aktuarialnego w ubezpieczeniach innych niż ubezpieczenia na życie w tym zakładzie).

Odpowiedzi:

....

Odp. a)

Formuła standardowa na SCR dla ryzyka składki rezerw jest następująca (zob. Art. 115 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$SCR_{nl\,prem\,res} = 3 \cdot \sigma_{nl} \cdot V_{nl}$$

gdzie σ_{nl} - odchylenie standardowe ryzyka składki i rezerw w ubezpieczeniach innych niż na życie wyznaczone zgodnie z art. 117 [ROZPORZĄDZENIE DELEGOWANE..., 2015]; V_{nl} - miara wielkości ryzyka składki i rezerw w ubezpieczeniach innych niż na życie, wyznaczone zgodnie z art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015]. Przy czym V_{nl} jest równe:

$$V_{nl} = \sum_s V_s,$$

gdzie V_s oznacza miarę wielkości ryzyka składki i rezerw dla segmentu s (wykaz segmentów podano w Aneksie II [ROZPORZĄDZENIE DELEGOWANE..., 2015].

Odchylenie standardowe σ_{nl} jest wyznaczane następująco (zob. Art. 117, Par.1 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$\sigma_{nl} = \frac{1}{V_{nl}} \cdot \sqrt{\sum_{s,t} CorrS_{s,t} \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t}$$

gdzie

$CorrS_{s,t}$- oznacza parametr zależności ryzyka składki i rezerw w ubezpieczeniach innych niż na życie dla segmentów s i t określony w Aneksie IV [ROZPORZĄDZENIE DELEGOWANE..., 2015]; σ_s, σ_t- odchylenia standardowe ryzyka składki i rezerw w ubezpieczeniach innych niż na życie odpowiednio dla segmentów s i t; V_s, V_t - miara wielkości ryzyka składki i rezerw odpowiednio dla segmentów s i t, o których mowa w art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015]. Odchylenie standardowe ryzyka składki i rezerw dla konkretnego segmentu s jest równe (zob. Art. 117, Par.2 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$\sigma_s = \frac{\sqrt{\sigma^2_{(pr,s)} \cdot V^2_{(pr,s)} + \sigma_{(pr,s)} \cdot V_{(pr,s)} \cdot \sigma_{(res,s)} \cdot V_{(res,s)} + \sigma^2_{(res,s)} \cdot V^2_{(res,s)}}}{V_{(pr,s)} + V_{(res,s)}}$$

gdzie: $\sigma_{(pr,s)}$ - odchylenie standardowe ryzyka składki dla segmentu s wyznaczone zgodnie z Art. 117, Par.3 [ROZPORZĄDZENIE DELEGOWANE..., 2015]); $\sigma_{(res,s)}$ - odchylenie standardowe ryzyka rezerw dla segmentu s określone w Aneksie II [ROZPORZĄDZENIE DELEGOWANE..., 2015]; $V_{(pr,s)}$, , $V_{(res,s)}$ – odpowiednio miara wielkości ryzyka składki i ryzyka rezerw dla segmentu s, o której mowa w art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015].

[ROZPORZĄDZENIE DELEGOWANE..., 2015] - ROZPORZĄDZENIE DELEGOWANE KOMISJI (UE) 2015/35 z dnia 10 października 2014 r. uzupełniające dyrektywę Parlamentu Europejskiego i Rady 2009/138/WE w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej i reasekuracyjnej (Wypłacalność II)

Stąd odchylenie standardowe ryzyka składki i rezerw:

- dla segmentu s jest równe: $\sigma_s = 0.082310388$
- dla segmentu t jest równe: $\sigma_t = 0.069282032$

Odchylenie standardowe σ_{nl} jest równe:

- $\sigma_{nl} = 0.065722201$- w przypadku $CorrS_{s,t} = 0.5$
- $\sigma_{nl} = 0.07579621$- w przypadku $CorrS_{s,t} = 1$

Na tej podstawie łączny wymóg kapitałowy dla ryzyka składki i rezerw rozważnych segmentów wynosi:

- $SCR_{nl\,prem\,res} = 78.867$- w przypadku $CorrS_{s,t} = 0.5$
- $SCR_{nl\,prem\,res} = 90.955$- w przypadku $CorrS_{s,t} = 1$

Wymóg kapitałowy dla ryzyka aktuarialnego w ubezpieczeniach innych niż ubezpieczenia na życie wyznaczamy stosując standardową formułę agregacji w Solvency II:

$$SCR = \sqrt{\sum_{i,j} Corr_{i,j} \cdot SCR_i \cdot SCR_j}$$

Korzystając z zapisu macierzowego otrzymujemy:

$$SCR = \sqrt{\begin{bmatrix} 78.867 & 100 & 10 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0,25 & 0 \\ 0,25 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 78.867 \\ 100 \\ 10 \end{bmatrix}} = 142.349$$

....

Odp. b)

Efekt dywersyfikacji (uzyskany w wyniku agregacji wymogów kapitałowych dla podmodułów ryzyka aktuarialnego w ubezpieczeniach innych niż ubezpieczenia na życie dla tego zakładu podmodułów):

- Bezwzględny: $D = 78.867 + 100 + 10 - 142.349 = 46.518$
- Względny: $d = \frac{46.518}{78.867+100+10} = 0.246$

Efekt dywersyfikacji (łączny, uwzgledniający także agregację ryzyka składki i rezerw analizowanych segmentów):

- Bezwzględny: $D = 90.955 + 100 + 10 - 142.349 = 58.606$
- Względny: $d = \frac{58.606}{90.955+100+10} = 0.292$

Uwaga! Oba rozwiązania były akceptowane.

Rozwiązanie:

Zadanie 7.

a) Scharakteryzować rodzinę rozkładów Tweediego (*Tweedie family of distributions*) i podać przykłady rozkładów należących do niej (co najmniej 4).
b) Omówić możliwość wykorzystania tej rodziny w taryfikacji.

Odpowiedzi:

....

Odp. a)

W odpowiedzi należało uwzględnić następujące elementy:

- Definicję rodziny rozkładów Tweediego: Są to rozkłady z wykładniczej rodziny rozkładów, dla których funkcja wariancji ma postać: $\vartheta(\mu) = \mu^p$.
- Przykłady rozkładów z tej rodziny (co najmniej 4):

Parametr p	Typ rozkładu	Nazwa	Wykorzystanie
$p < 0$	Ciągła	-	-
$p = 0$	Ciągła	Normalny	-
$0 < p < 1$	Nie istnieje	-	-
$p = 1$	Dyskretny	Poissona	Liczba szkód
$1 < p < 2$	Mieszany, nieujemny	Złożony Poissona	Składka czysta: *szkody/ekspozycja*
$p = 2$	Ciągła, dodatnia	Gamma	Wysokość poj. szkody
$2 < p < 3$	Ciągła, dodatnia	-	Wysokość poj. szkody
$p = 3$	Ciągła, dodatnia	Odwrotny Gaussa	Wysokość poj. szkody
$p > 3$	Ciągła, dodatnia	-	Wysokość poj. szkody

....

Odp. b)

W odpowiedzi należało uwzględnić następujące elementy:

- Wskazanie, że rodzinę można wykorzystać w taryfikacji jednomodelowej. Omówienie wad i zalet tej metody taryfikacji.
- Wskazanie, że rodzinę można wykorzystać do modelowania liczby szkód.
- Wskazanie, że rodzinę można wykorzystać do modelowania wysokości pojedynczej szkody.

Rozwiązanie:

Zadanie 8.

Założono, że obserwacje $y_1, y_2, \dots, y_n$ mają rozkład wykładniczy:

$$f(y_i) = \frac{1}{\mu_i} \exp\left(-\frac{1}{\mu_i} y_i\right), i = 1,2, \dots, n.$$

Dopasowano do nich uogólniony model liniowy o następującej specyfikacji:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}.$$

Poniżej podano odpowiednio wektor oszacowanych parametrów $\hat{\beta}$ oraz 6 pierwszych wierszy wektora obserwacji Y, macierzy modelu X oraz macierz daszkowej H:

$$\hat{\beta} = \begin{bmatrix} 6.0691 \\ -0.3923 \\ -0.0415 \\ -0.2886 \\ 0.2075 \end{bmatrix}, \qquad Y = \begin{bmatrix} 80 \\ 13 \\ 55 \\ 161 \\ 49 \\ 33 \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & 4 & 10 & 1 & 0 \\ 1 & 5 & 20 & 0 & 0 \\ 1 & 2 & 10 & 1 & 1 \\ 1 & 3 & 15 & 0 & 1 \\ 1 & 4 & 20 & 0 & 0 \\ 1 & 2 & 10 & 0 & 1 \end{bmatrix}$$

$$H = \begin{vmatrix} 0,4890 & 0,1356 & 0,1852 & -0,0608 & 0,0841 & -0,0049 \\ 0,1356 & 0,3041 & -0,0909 & 0,1002 & 0,2112 & 0,0300 \\ 0,1852 & -0,0909 & 0,3653 & 0,1497 & -0,0446 & 0,2566 \\ -0,0608 & 0,1002 & 0,1497 & 0,3379 & 0,0946 & 0,3596 \\ 0,0841 & 0,2112 & -0,0446 & 0,0946 & 0,2087 & 0,1043 \\ -0,0049 & 0,0300 & 0,2566 & 0,3596 & 0,1043 & 0,5019 \end{vmatrix}$$

Dla drugiej obserwacji wyznaczyć:

a) resztę Pearsona,

b) miarę Cooka (*Cook's distance*).

Odpowiedzi:

....

Odp. a)

1. Reszta Pearsona: $e_{P,i} = \frac{y_i - \hat{\mu}_i}{\sqrt{D^2(y_i|\boldsymbol{x})/\phi}}$.

W modelu wykładniczym parametr dyspersji $\phi = 1$ oraz $D^2(y_i|\boldsymbol{x}) = \mu_i^2$.

Wartość teoretyczna dla obserwacji nr 2 wynosi:

$\hat{\mu}_2 = \exp(3.2776) = 26.51206732.$

Stąd:

$$e_{P,2} = \frac{13 - 26.51206732}{26.51206732} = -0,50965725$$

....

Odp. b)

Miara Cooka: $D_i = \frac{e_{PS,i}^2}{k} \cdot \frac{h_i}{1-h_i}$,

gdzie:

$e_{PS,i} = \frac{y_i - \widehat{\mu}_i}{\sqrt{D^2(y_i|\boldsymbol{x})(1-h_i)}}$ - standaryzowana reszta Pearsona;

k – liczba parametrów modelu;

h_i – i-ty element leżący na głównej przekątnej macierzy „daszkowej" H.

Zatem dla szóstej obserwacji:

$e_{PS,2} = -0.610948761$

skąd $D_2 = 0.032621893$

Rozwiązanie:

Zadanie 9.

Straty

- dla pierwszej linii biznesu (LOB1) są modelowane za pomocą zmiennej losowej X_1 o rozkładzie normalnym z parametrami $E(X_1) = 0$ i $D(X_1) = 100$,
- dla drugiej linii biznesu (LOB2) - zmiennej losowej X_2 o rozkładzie normalnym z parametrami $E(X_2) = 0$ i $D(X_2) = 200$.

Łączne wymogi kapitałowe dla tych dwóch linii (czyli dla $S = X_1 + X_2$) wyznaczono z wykorzystaniem dwóch następujących metod:

- **Metoda M1.** Zastosowano standardową metodę wariancji-kowariancji wykorzystywaną w Solvency II. Przyjęto współczynnik korelacji liniowej między X_1 i X_2 równy 0.60.
- **Metoda M2.** Najpierw zidentyfikowano strukturę zależności między stratami X_1 i X_2. Okazało się, że najlepiej opisuje ją kopula Gumbela z parametrem $\theta = 2.5$. Następnie wykorzystując metodę symulacji wyznaczono wymóg kapitałowy na podstawie następującej formuły: $\kappa = VaR_{0.995}(S) - E(S)$. Wartość zagrożoną $VaR_{0.995}(S)$ uzyskano na podstawie podanych w poniższej tabeli wartości statystyk pozycyjnych $S_{(1081:1000)}, \ldots, S_{(1100:1100)}$ dla 1100 wartości zmiennej S uzyskanych w drodze symulacji. Wartości $\boldsymbol{B}$ i $\boldsymbol{C}$ odpowiadają następującym parom: $(u_B, v_B) = (0.985 \quad 0.991)$ i $(u_C, v_C) = (0.989 \quad 0.990)$, które zostały wygenerowane z zastosowanej kopuli.

1081	*1082*	*1083*	*1084*	*1085*	*1086*	*1087*	*1088*	*1089*	*1090*
562,21	564,97	589,17	591,50	591,56	592,31	606,85	619,17	626,53	628,97

1091	*1092*	*1093*	*1094*	*1095*	*1096*	*1097*	*1098*	*1099*	*1100*
643,25	674,91	682,24	***B***	***C***	700,02	768,78	845,45	852,36	864,70

a) Wyznaczyć różnicę między łącznymi wymogami kapitałowymi dla analizowanych linii biznesu wyznaczonymi metodą **M1** i **M2**.

b) Wyznaczyć różnicę między efektami dywersyfikacji uzyskanymi w wyniku zastosowania metod **M1** i **M2**.

Odpowiedzi:

....

Odp. a)

Metoda **M1**

Wymóg dla LOB1: $VaR_{0.995}(LOB1) = \Phi^{-1}(0.995) \cdot 100 = 2.576 \cdot 100 = 257.6$

Wymóg dla LOB1: $VaR_{0.995}(LOB2) = \Phi^{-1}(0.995) \cdot 200 = 2.576 \cdot 200 = 515.2$

Łączny wymóg: $\kappa^{(M1)} = \sqrt{[257.6 \quad 515.2] \cdot \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \cdot \begin{bmatrix} 257.6 \\ 515.2 \end{bmatrix}} = 700.748$

Metoda **M2**

$\boldsymbol{B} = F_1^{-1}(0.985) + F_2^{-1}(0.991) = \Phi^{-1}(0.985) \cdot 100 + \Phi^{-1}(0.991) \cdot 200 =$
$= 2.170 \cdot 100 + 2.366 \cdot 200 = 690.2$

$\boldsymbol{C} = F_1^{-1}(0.989) + F_2^{-1}(0.990) = \Phi^{-1}(0.989) \cdot 100 + \Phi^{-1}(0.990) \cdot 200 =$
$= 2.290 \cdot 100 + 2.326 \cdot 200 = 694.2$

W celu oszacowania VaR-u korzystamy ze wzoru:

$$\widehat{VaR}_\alpha(X) = X_{[n \cdot \alpha]+1:n}\,.$$

Stąd otrzymujemy

$$VaR_{0.995}(S) = S_{[1100 \cdot 0.995]+1:1100} = S_{1095:1100} = 694.2$$

i łączny wymóg

$$\kappa^{(M2)} = VaR_{0.995}(S) - E(S) = 694.2 - 0 = 694.2$$

Różnica: $\kappa^{(M1)} - \kappa^{(M2)} = 6.548$

....

Odp. b)

Efekt dywersyfikacji dla **M1**:

- Bezwzględny: $D = 257.6 + 515.2 - 700.748 = 772.8 - 700.748 = 72.052$
- Względny: $d = \frac{72.052}{772.8} = 0.093$

Efekt dywersyfikacji dla **M2**:

- Bezwzględny: $D = 257.6 + 515.2 - 694.2 = 772.8 - 694.2 = 78.6$
- Względny: $d = \frac{78.6}{772.8} = 0.102$

Różnica:

- Bezwzględny: 6.548
- Względny: 0.009

Rozwiązanie:

Zadanie 10.
Poniżej podano następujące informacje związane z liniowym modelem regresji postaci $Y \sim X_1 + X_2 + X_3$.

- Tabela analizy wariancji modelu liniowego: $X_1 \sim X_2 + X_3$

Response: X_1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X_2	1	226882	226882	1376.5863	< 2e-16
X_3	1	847	847	5.1408	0.02386
Residuals	434	71530	165		

- Tabela analizy wariancji modelu liniowego: $X_2 \sim X_1 + X_3$

Response: X_2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X_1	1	326.08	326.08	1364.2778	<2e-16
X_3	1	0.29	0.29	1.2143	0.2711
Residuals	434	103.73	0.24		

- Tabela analizy wariancji modelu liniowego: $X_3 \sim X_1 + X_2$

Response: X_3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X_1	1	65.3	65.260	7.0723	0.008118
X_2	1	11.2	11.205	1.2143	0.271091
Residuals	434	4004.8	9.228		

a) Omówić problem współliniowości zmiennych objaśniających w liniowym modelu regresji. Zaproponować sposoby „radzenia sobie" z tym problemem (gdy wystąpi).
b) Na podstawie powyższych informacji wyznaczyć współczynnik inflacji wariancji (współczynnik VIF – *Variance Inflation Factor*) dla każdej ze zmiennych objaśniających.
c) Czy wystąpił problem współliniowości w przedstawionym w zadaniu modelu?

Odpowiedzi:

....

Odp. a)
W odpowiedzi należało uwzględnić następujące elementy:

- Opis problemu. W modelu regresji liniowej z wieloma zmiennymi objaśniającymi zakłada się, że nie występują zależności liniowe między zmiennymi niezależnymi. Oznacza to, że macierz $X'X$ jest pełnego rzędu i istnieje macierz do niej odwrotna. Problem współliniowości występuje, gdy
    - co najmniej jedna zmienna objaśniająca jest liniowo zależna od jednej lub kilku pozostałych. Wtedy $X'X$ nie jest macierzą pełnego rzędu, nie jest więc odwracalna i nie można wyznaczyć ocen parametrów modelu.
    - występuje silne skorelowanie zmiennych objaśniających. Wtedy można wyznaczyć oceny parametrów, ale mogą one być niewiarygodne.

  W modelowaniu najczęściej mamy do czynienia z drugim z wyżej wymienionych przypadków.

- Wskazanie oznak możliwej współliniowości, np. bardzo wysoki współczynnik determinacji i nieistotne parametry.
- Wskazanie sposobów „radzenia sobie" z tym problemem. Np.
    - usunięcie wybranych (z najbardziej skorelowanych) zmiennych, tak aby zmniejszył się VIF pozostałych,
    - wprowadzenie ograniczeń na parametry (regresja grzbietowa),
    - skonstruowanie nowych zmiennych z użyciem np. PCA.

....

Odp. b)

Współczynnik inflacji wariancji

$$VIF_j = \frac{1}{1 - R_j^2}$$

Obliczenia dla VIF_1:

$$R_1^2 = \frac{226882 + 847}{226882 + 847 + 71530} = \frac{227729}{299259} = 0.761$$

$$VIF_1 = \frac{1}{1 - R_1^2} = \frac{1}{1 - 0.761} = 4.184$$

Analogicznie otrzymujemy:

$VIF_2 = 4.1463$

$VIF_1 = 1.0191$

....

Odp. c)

Przyjmuje się, że problem współliniowości występuje, gdy VIF jest większy od 10. W literaturze można spotkać także wartość graniczną równą 5.

W analizowanym przypadku nie występuje.

Rozwiązanie:

Sesja egzaminacyjna w dniu 5 marca 2019 r.

Modelowanie

Arkusz ocen

Zadanie nr	Punktacja
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	